BMC
Psychiatry

RESEARCH ARTICLE   **Open Access**

Psychometric properties of responses by clinicians and older adults to a 6-item Hebrew version of the Hamilton Depression Rating Scale (HAM-D$_6$)

Yaacov G Bachner[1*], Norm O'Rourke[2], Margalit Goldfracht[3], Per Bech[4] and Liat Ayalon[5]

Abstract

Background: The Hamilton Depression Rating Scale (HAM-D) is commonly used as a screening instrument, as a continuous measure of change in depressive symptoms over time, and as a means to compare the relative efficacy of treatments. Among several abridged versions, the 6-item HAM-D$_6$ is used most widely in large degree because of its good psychometric properties. The current study compares both self-report and clinician-rated versions of the Hebrew version of this scale.

Methods: A total of 153 Israelis 75 years of age on average participated in this study. The HAM-D$_6$ was examined using confirmatory factor analytic (CFA) models separately for both patient and clinician responses.

Results: Reponses to the HAM-D$_6$ suggest that this instrument measures a unidimensional construct with each of the scales' six items contributing significantly to the measurement. Comparisons between self-report and clinician versions indicate that responses do not significantly differ for 4 of the 6 items. Moreover, 100% sensitivity (and 91% specificity) was found between patient HAM-D$_6$ responses and clinician diagnoses of depression.

Conclusion: These results indicate that the Hebrew HAM-D$_6$ can be used to measure and screen for depressive symptoms among elderly patients.

Keywords: Depression, Hamilton depression rating scale, Hebrew, Elderly

Background

Depression is a common debilitative psychiatric condition ranked high in prevalence among all mental health conditions [1]. Lifetime prevalence may be as high as 20% [2] and, at any one time, 5–10% of the world's population meets diagnostic criteria for a major depressive episode [3]. Depression is projected to be the second leading cause of disability worldwide in 2020 [4].

Clinical depression is common in primary care with rates of prevalence among older adults ranging between 4–24% [5,6]. Untreated elderly patients are at higher risk of morbidity and mortality [7] and experience slower rates of recovery [6,8]. Moreover, chronic depression is a significant risk factor for dementia [9].

Given that depression is amenable to treatment, valid and reliable screening tools are necessary to identify this patient population. Among existing instruments, the clinician-administered Hamilton Depression Rating Scale (HAM-D) was first developed to assess the efficacy of the first generation of antidepressant medications [10]; the HAM-D has since become the gold standard for measuring symptom severity and change in randomized clinical trials. Among various formats (17, 21, 24 & 28 items) [10,11], the 17-item (HAM-D$_{17}$) has been used most frequently. Scale items measure mood, insomnia, anhedonia, agitation, gastro-intestinal and other somatic symptoms, weight change, suicidal ideation, hypochondriasis, anosognosia, and psychomotor and cognitive retardation.

* Correspondence: Bachner@bgu.ac.il
[1]Department of Public Health and the Center for Multidisciplinary Research in Aging, Faculty of Health Sciences, Ben-Gurion University of the Negev, P.O. B 653, Beer-Sheva 84105, Israel
Full list of author information is available at the end of the article

Despite widespread usage, various researchers have questioned whether the HAM-D$_{17}$ is a unidimensional or multidimensional instrument [12-15]. This is problematic as multi-factorial measurement may impede the detection of symptom change over time, treatment response characteristics [16] and the ability to distinguish the relative efficacy of treatments [13]. This assertion is supported by meta-analytic study findings indicating that certain scale items are less sensitive to measurement of symptom severity. In addition, some items have comparatively poor inter-rater and retest reliability, and the response-option format may not be optimal [17]. In light of these findings, some have suggested that the 17-item HAM-D may be less than ideal for clinical research applications [14,15,17,18].

These limitations have led researchers to propose abridged versions of the HAM-D that are quick to administer yet sensitive to measurement of symptom levels, change over time and relative differences in treatment efficacy. For instance, Maier and Philipp [19] proposed a 6-item version of the HAM-D. More recently, an 8-item version was devised by Gibbons and colleagues [20] by applying item response theory. Research to date suggests that both versions are sensitive to change over time and can identify patients in remission [21,22]. Recently, a scale consisting of 7 items was also suggested [23]. The items were empirically identified on the basis of response frequency and sensitivity to change of the individual HAM-D items with depressed samples [24].

Among the abridged versions of the Hamilton scale, the most frequently used was developed by Bech et al. (HAM-D$_6$) [25]. Using item analysis, these researchers [25] have proposed a 6-item HAM-D as a unidimensional measure of depressive symptomatology [14]. This HAM-D$_6$ is composed of items measuring core symptoms of depression (i.e., depressed mood, self-esteem and feelings of guilt, social interaction and interests, psychomotor retardation, anxiety, and somatic symptoms). Compared to the HAM-D$_{17}$, this assessment appears to measure a unidimensional construct [13-15,17,25,26], and it is as sensitive [14] or more sensitive in detecting drug–placebo or drug–drug differences [27,28]. The authors of a recent study with older adults that compared six depression scales concluded that the HAM-D$_6$ was the only one to demonstrate total scalability, and that it had the greatest external validity [18].

This scale. may be especially appropriate for use by both older persons and clinicians; its relative brevity makes it comparatively easy for older persons to complete and clinicians to administer. However, to the best of our knowledge, the psychometric properties of responses to the Hebrew HAM-D$_6$ had yet to be examined. Thus, the current study examined and compared self-report and clinician responses to the Hebrew HAM-D$_6$ for elderly patients.

Methods

Scale translation

The HAM-D$_6$ was first translated from English to Hebrew by a bilingual psychologist, in keeping with accepted procedures [29]. The translated version was back translated and modified until it was comparable to the original version.

Training procedures

Two graduate research assistants completed a three-day training course in the administration of study measures. After watching a training tape and receiving instructions, they administered study measures in mock interviews until acceptable inter-rater reliability was established vis-à-vis semi-structured clinical assessments. Research assistants' HAM-D$_6$ scores did not significantly differ from corresponding patient HAM-D responses suggesting no discernible between-rater differences, χ^2 ($df = 1$) $= 1.31$, $p = .25$.

Recruitment

Participants were recruited in the waiting rooms of two primary care clinics operated by Clalit Health Services (Israel's largest health insurance provided serving 53% of the population). One clinic is located in the center and the other in the north of Israel (Tel Aviv and Haifa, respectively). Inclusion criteria were: 60+ years of age, fluent in Hebrew, and no pronounced cognitive loss (determined using a 6-item screening measure [30]). Participant recruitment took place between May, 2008 and February, 2009.

Research assistants approached patients to request their participation in this study. Participation was voluntary and no remuneration was provided. Those who took part provided written consent. This study was approved by the Helsinki Committee of the Clalit Health Care Services.
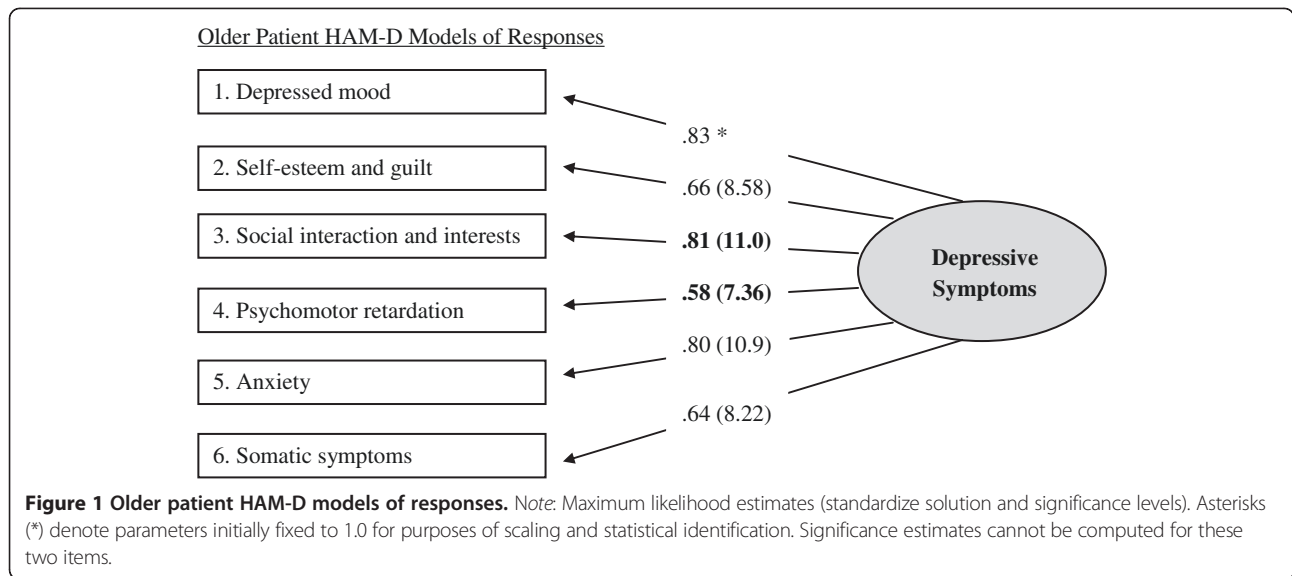
Measures

The Structured Clinical Interview for DSM-IV (SCID-I)

The SCID-I is a semi-structured interview to assist clinicians in making a DSM-IV Axis I diagnosis [31]. Only those modules pertaining to depression and dysthymia were administered in the present study. The Hebrew version of the SCID-I was translated and validated by Shalev et al. [32]. All study participants were interviewed using this instrument.

The 6-item Hamilton (HAM-D6)

The self- and clinician-administered versions of the HAM-D$_6$ measure depressed mood, self-esteem and guilt, social interaction and interests, psychomotor retardation, anxiety, and somatic symptoms. Items are provided along 5-point scales, with the exception of the

Older Patient HAM-D Models of Responses

Figure 1 Older patient HAM-D models of responses. Note: Maximum likelihood estimates (standardize solution and significance levels). Asterisks (*) denote parameters initially fixed to 1.0 for purposes of scaling and statistical identification. Significance estimates cannot be computed for these two items.

somatic symptoms item (where responses were provided on a 3-point scale). As a screening measure, scores of 7+ suggest clinically significant depressive symptomatology [33]. Whereas the self-report HAM-D_6 is based solely on patient responses, the clinician-administered version integrates patients' responses and clinical observation.
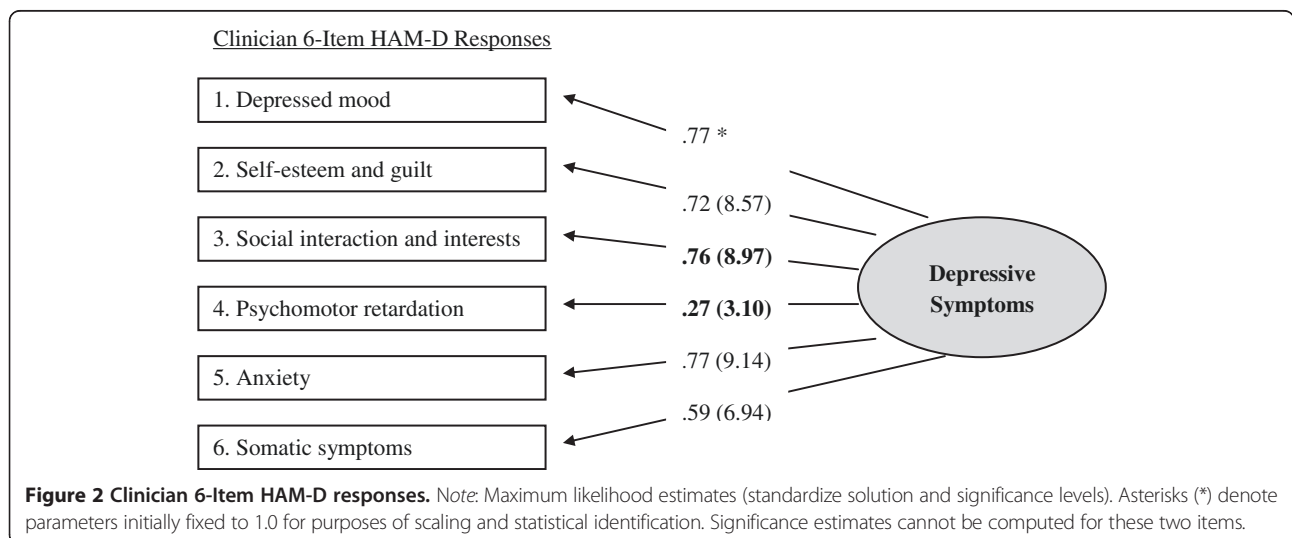
Analytic strategy

We set out to ascertain if the HAM-D_6 measures a unidimensional construct, as proposed by Bech et al. [25]. This hypothesis was tested using confirmatory factor analyses. Both self- and clinician-administered versions of the HAM-D_6 were next compared to assess the relative contribution of items to measurement (invariance or equivalence analyses). Subsequent analyses were undertaken comparing responses for each patient

(self and corresponding clinician HAM-D_6 responses). Comparisons between SCID diagnoses of a major depressive episode and the patient HAM-D_6 responses were made to estimate sensitivity and specificity of the scale. Lastly, item-level analyses were computed (intraclass correlation coefficients) to determine if there was agreement between patients and their clinicians for each item.

Results

This sample was composed of 153 patients 75 years of age on average (range 59–98; SD = 8.1). The majority of participants were male (91/153 or 59.5%). Eighty seven (56.9%) were currently married and living with a spouse, 54 (35.3%) were widowed, and 12 (8.8%) were divorced or lived alone. Respondents' mean level of education was

Clinician 6-Item HAM-D Responses

Figure 2 Clinician 6-Item HAM-D responses. Note: Maximum likelihood estimates (standardize solution and significance levels). Asterisks (*) denote parameters initially fixed to 1.0 for purposes of scaling and statistical identification. Significance estimates cannot be computed for these two items.

Table 1 Invariance analyses of older patient and clinician 6-Item HAM-D responses

Model	df	χ^2	Δdf	$\Delta \chi^2$	AGFI	SRMR	CFI	RMSEA (RMSEA-CL$_{90}$)
1. Baseline	17	30.745	–	–	.93	.0352	.98	.052 (.020 – .081)
2. Self-esteem and guilt	18	31.901	1	1.156	.93	.0348	.98	.051 (.019 – .079)
3. Social interaction and interests	19	36.796	1	4.896 **	.92	.0348	.97	.056 (.028 – .083)
4. Psychomotor retardation	20	67.602	1	30.806 **	.86	.0746	.93	.089 (.066 – .113)
5. Anxiety	21	69.090	1	1.488	.87	.0756	.93	.087 (.065 – .110)
6. Somatic symptoms	22	69.127	1	0.037	.87	.0759	.93	.084 (.062 – .107)

Note. df = degrees of freedom, AGFI = Adjusted Goodness of Fit Index, SRMR = Standardized Root Mean Residual, CFI = Comparative Fit Index, RMSEA = Root Mean Square Error of Approximation, CL$_{90}$ = 90% Confidence Limits.

11.8 years (range 4–20; SD = 3.1), and the majority (63.4%) ranked their economic status as fair.

HAM-D$_6$ as a screening measure

As previously mentioned, Bech et al. [33] suggest that a HAM-D$_6$ score of 7+ is suggestive of clinically significant depressive symptoms (i.e., warranting thorough clinical assessment). Comparing patient and clinician ratings, agreement as calculated using the kappa coefficient was in fair range (k = .26; [34]). Where there was a discrepancy between the two, 13 patients provided responses in clinical range, whereas physicians' responses indicated these patients were euthymic. A similar finding emerged comparing patient HAM-D$_6$ responses with SCID diagnoses of a current major depressive episode (k = .20; linear weighted). Where there was a discrepancy, 14 patients provided HAM-D$_6$ responses in clinical range, while the SCID diagnoses indicated no major depressive episode. However, these percentages indicate 100% sensitivity for the patient version of the HAM-D$_6$ (true positives) and 91% specificity (true negatives).

Confirmatory factor analytic models

Confirmatory factor analytic (CFA) models were computed separately for older patients ($\chi^2[df = 7] = 23.80$, $p < .01$) and corresponding clinician HAM-D$_6$ scores, ($\chi^2[df = 9] = 16.93$, $p = .05$). Goodness of fit indices for both models were within optimal parameters [35].

Table 2 Intra-class correlation coefficients between older patient and clinician HAM-D$_6$ responses

Item	ICC
1. Depressed mood	.78
2. Self-esteem and guilt	.73
3. Social interaction and interests	.74
4. Psychomotor retardation	.12
5. Anxiety	.62
6. Somatic symptoms	.64

Moreover, each of the six items contributed significantly to measurement of a single higher-order construct (i.e., all item *t* values > 1.96); see Figures 1 and 2. For both patient and clinician versions, the HAM-D$_6$ appears to measure a unidimensional depression construct.

Next, invariance analyses were undertaken to compare solutions between CFA models. These analyses indicated that responses did not significantly differ for 4 of 6 items. However, responses for the *social interaction and interests* and *psychomotor retardation* items did differ. Both contributed to measurement of depression as reported by patients to a greater degree than that reported by the clinicians. See Table 1.

Intra-class correlation coefficients

Intra-class correlation coefficients (ICC) were next computed to directly compare HAM-D$_6$ ratings for patient–clinician pairings (i.e., patient self-report vs. corresponding clinician ratings for that patient). ICC values were within adequate parameters for items 1–3 (depressed mood, self-esteem and guilt, social interaction and interests), low for items 5–6 (anxiety, somatic symptoms), but very low for item 4 (psychomotor retardation). This is consistent with invariance analyses reported above, see Table 2.

Discussion

The goal of this study was to assess the psychometric properties of self-report vs. clinician versions on the Hebrew HAM-D$_6$. Results indicated that each of the six scale items contributed significantly to the measurement (both for patients and clinicians) and that HAM-D$_6$ responses indeed measure a single depression construct. These findings are in accord with previously reported findings [13-15,25,26,33].

Comparing clinician and patient HAM-D$_6$ responses indicate satisfactory correspondence between the two. Moreover, when patient HAM-D$_6$ responses were compared to SCID diagnoses of major depressive episodes, sensitivity and specificity were measured as 100% and 91%, respectively.

These findings suggest that a 7+ HAM-D$_6$ score is an effective threshold value. Most notably, responses by older adults, themselves, enable effective depression screening between euthymic patients and those reporting pronounced depressive symptomatology.

In addition, findings indicate that responses do not differ significantly for 4 of the 6 items suggesting that patients and clinicians appear to interpret and respond to these HAM-D$_6$ items in a consistent manner. Furthermore, the intra-class correlations for 5 of the 6 items were found to be above 0.60. This congruence between patients and clinicians for most scale items implies that patients' responses can be trusted and accepted as a valid evaluation of depression.

Responses do differ, however, for the *social interaction and interests* and *psychomotor retardation* items. For both items, patients' responses contributed more to the measurement of depression than clinicians' responses. Furthermore, the intra-correlation coefficient for the *psychomotor retardation* was found to be very low, but for the *social interaction and interests* item, an adequate correlation emerged.

In light of these intriguing results, we re-examined the Hebrew translations in order to ascertain where refinements are warranted. In English, the second response option for the *social interaction and interests* item reads: "I have felt that I have had difficulty performing my daily activities, but I was still able to perform them with great effort." The current Hebrew wording translates to: "I had difficulty performing my daily activities, but I was still able to perform routine activities".

The fourth response of this item in English reads: "I have not been able to do any of the simplest day-to-day activities without help," and the current Hebrew wording translates to: "I have not been able to do any of the simple day-to-day activities without help." Although the difference appears minimal, it might have had an effect on the results.

In English, the third and fourth response options for the *psychomotor retardation* item reads: "I have felt clearly slowed down or subdued or have been talking much less than usual," and "I have hardly been talking at all or feel extremely slowed down at the time." The corresponding Hebrew wording translates to: "I have felt clearly slowed down or passive and have been talking much less than usual," and "I have hardly been talking at all and feel extremely slowed down all the time." We recommend that corrections in translation be made for future studies using the self-report Hebrew HAM-D$_6$.

Several limitations of the study need to be acknowledged: a) we do not have data on non-participants and cannot compare this group to our sample, b) we do not have medication data for this sample, c) this is a relatively small sample size, and d) the research assistants that assessed the participants SCID were aware of their HAM-D$_6$ scores. Therefore, future studies need to examine the Hebrew HAM-D$_6$ with larger samples of participants from different age groups derived by random recruitment.

Conclusion

Nonetheless, in the light of our results, the Hebrew HAM-D$_6$ can be used to measure and screen depressive symptoms among elderly persons. Future psychometric research is required to ascertain whether the above suggested revisions will further improve the psychometric properties of responses to this Hebrew version of the HAM-D$_6$.

Authors' contribution
YGB wrote the manuscript and made critical revisions. LA, MG and PB conceived, developed and designed the study. LA also supervised the data collection. NO'R carried out the data analysis, wrote the results section and made critical revisions. All authors have read and approved the final manuscript.

Author details
[1]Department of Public Health and the Center for Multidisciplinary Research in Aging, Faculty of Health Sciences, Ben-Gurion University of the Negev, P.O. B 653, Beer-Sheva 84105, Israel. [2]Faculty of Arts and Social Sciences, Simon Fraser University, Burnaby, (BC), Canada. [3]Community Division, Clalit Health Services, Tel Aviv, Israel and Department of Family, Health Care, Bruce Rappaport Faculty of Medicine, The Technion, Haifa, Israel. [4]Department of Psychiatry, Frederiksborg General Hospital, Hilleord, Denmark. [5]Louis and Gabi Weisfeld School of Social Work, Bar-Ilan University, Ramat Gan, Israel.

References
1. Richards D: **Prevalence and clinical course of depression: a review.** *Clin Psychol Rev* 2011, **31**(7):1117–1125.
2. American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders.* Washington, DC: Revised 4th ed; 2000.
3. Moussavi S, Chatterji S, Verdes E, Tandon A, Patel V, Ustun B: **Depression, chronic diseases, and decrements in health: results from the World Health Surveys.** *Lancet* 2007, **370**:851–858.
4. Murray CJ, Lopez AD: **Global mortality, disability, and the contribution of risk factors: Global Burden of Disease Study.** *Lancet* 1997, **349**:1436–1442.
5. Van Marwijk H, Hoeksema HIL, Hermas J, Kaptein AA, Mulder JD: **Prevalence of depressive symptoms and depressive disorder in primary care patients over 65 years of age.** *Fam Pract* 1994, **11**:80–84.
6. Williams JWJ, Kerber CA, Mulrow CD, Medina A, Aguilar C: **Depressive disorders in primary care: prevalence, functional disability, and identification.** *J Gen Intern Med* 1995, **10**:7–12.
7. Cuijpers F, Smith P: **Excess mortality in depression: a meta-analysis of community studies.** *J Affect Disord* 2002, **72**:36–227.
8. Kiecolt-Glaser JK, Glaser R: **Depression and immune function: central pathways to morbidity and mortality.** *J Psychosom Res* 2002, **53**:873–876.
9. Saczynski JS, Beiser A, Seshadri S, Auerbach S, Wolf PA, Au R: **Depressive symptoms and risk of dementia: The Framingham Heart Study.** *Neurology* 2010, **75**:35–41.
10. Hamilton M: **A rating scale for depression.** *J Neurosurg* 1960, **23**:56–62.

11. Hamilton M: **Development of a rating sale for primary depressive illness.** *Br J Soc Clin Psychol* 1967, **6**:278–296.
12. Bech P, Allerup P, Gram LFN, Rosenberg R, Jacobsen O, Nagy A: **The Hamilton Depression Scale: evaluation of objectivity using logistic models.** *Acta Psychiatr Scand* 1981, **63**:290–299.
13. Carmody TJ: **The Montgomery–Asberg and the Hamilton ratings of depression: a comparison of measures.** *Eur Neuropsychopharmacol* 2006, **16**:601–611.
14. Lecrubier Y, Bech P: **The Ham D6 is more homogeneous and as sensitive as the Ham D17.** *Eur Psychiat* 2007, **22**:252–255.
15. Licht RW, Qvitzau S, Allerup P, et al: **Validation of the Bech-Rafaelsen Melnacholia Scale and the Hamilton Depression Scale in patients with major depression: Is the total score a valid measure of illness severity?** *Acta Psychiatr Scand* 2005, **111**:144–149.
16. Santor DA, Coyne JC: **Examining symptoms expression as a function of symptom severity: item performance on the Hamilton Rating Scale for depression.** *Psychol Assessment* 2001, **13**:127–139.
17. Bagby RM, Ryder AG, Schuller DR, Marshall MB: **The Hamilton Depression Rating Scale: Has the gold standard become a lead weight?** *Am J Psychiatry* 2004, **161**:2163–2177.
18. Korner A, Lauritzen L, Abelskov K, et al: **Ratings scales for depression in the elderly: external and internal validity.** *J Clin Psychiatry* 2007, **68**:384–389.
19. Maier W, Philipp M: **Improving the assessment of severity of depressive states: a reduction of the Hamilton Depression Scale.** *Pharmacopsychiatry* 1985, **18**:114–115.
20. Gibbons RD, Clark DC, Kupfer DJ: **Exactly what does the Hamilton Depression Rating Scale measure?** *J Psychiatr Res* 1993, **27**:259–273.
21. Entsuah R, Shaffer M, Zhang J: **A critical examination of the sensitivity of unidimensional scales derived from the Hamilton Depression Rating Scale of antidepressant drug effects.** *J Psychiatr Res* 2002, **36**:437–448.
22. Faries D, Herrera J, Rayamajhi J, DeBrota D, Demitrack M, Potter WZ: **The responsiveness of the Hamilton Depression Rating Scale.** *J Psychiatr Res* 2000, **34**:3–10.
23. McIntyre RS, Konarski JZ, Mancini DA, Fulton KA, Parikh SV, Grigoriadis S, Grupp LA, Bakish D, Filteau M, Gorman C, Nemeroff CB, Kennedy SH: **Measuring the severity of depression and remission in primary care: validation of the HAMD-7 scale.** *CMAJ* 2005, **173**:1327–1334.
24. Ballesteros J, Bobes J, Bulbena A, Luque A, Dal-Ré R, Ibarra N, Güemes I: **Sensitivity to change, discriminative performance, and cutoff criteria to define remission for embedded short scales of the Hamilton Depression Rating Scale (HAMD).** *J Affect Disord* 2007, **102**:93–99.
25. Bech P, Gram LF, Dein E, Jacobson O, Vitger J, Bolwing TG: **Quantitative rating of depressive states.** *Acta Psychiatr Scand* 1975, **51**:161–170.
26. Bech P, Wilson BP, Wessel T, Junde M, Fava M: **A validation analysis of self-reported HAM-D6 versions.** *Acta Psychiatr Scand* 2009, **119**:298–03.
27. Bech P, Cialdella P, Haugh MC, et al: **Meta-analysis of randomized controlled trials of fluoxetine v. placebo and tricyclic anidepressants in the short-term treatment of major depression.** *Br J Psychiatry* 2000, **176**:421–428.
28. Faries D, Herrera J, Raymajhi J, DeBrota D, Demitrack M, Potter WZ: **The responsiveness of the Hamilton Depression Rating Scale.** *J Psychiatr Res* 2000, **34**:3–10.
29. Koller M, Aaronson NK, Blazeby J, et al: **Translation procedures for standardized quality of life questionnaires: The European Organization for Research and Treatment of Cancer (EORTC) approach.** *Eur J Cancer* 2007, **43**:1810–1820.
30. Callahan EJ, Bertakis KD, Azari R, Robbins JA, Helms LJ, Leigh JP: **Association of higher costs with symptoms and diagnosis of depression.** *J Fam Pract* 2002, **51**:540–544.
31. First MB, Spitzer RI, Gibbon M, Williams JBW: *Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I).* Administration Booklet: Clinician Version; 1997.
32. Shalev A, Sahar T, Abramovitz M: *Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I).* Department of Psychiatry: Hadassah University Hospital, Jerusalem, Israel; 1996.
33. Bech P, Lunde M, Bech-Andersen G, Lindberg L, Martiny K: **Psychiatric outcome studies: Does treatment help the patient?** *Nord J Psychiatry* 2007, **61**(46):4–80.
34. Landis JR, Koch GG: **The measurement of observer agreement for categorical data.** *Biometrics* 1977, **33**:159–174.
35. Hu LT, Bentler PM: **Cut off criteria for fit indices in covariance structure analysis: conventional criteria versus new alternatives.** *Struct Equ Modeling* 1999, **6**:1–55.